Improving NASA Earth science data and information access through natural language processing based data analysis and visualization

Zhong Liu[1,2], Shimei Pan[3], Jianwu Wang[3], Angela Li[1], and David Meyer[1]

[1]NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC)
[2]CSISS, George Mason University
[3]Dept. of Information Systems, UMBC

2018 NASA Goddard Workshop on Artificial Intelligence

Motivation

NASA:

- The Research Access initiative is part of the agency's framework for increasing public access to scientific publications and digital scientific data.
- The initiative follows the release of White House Office of Science and Technology Policy's (OSTP) memorandum "Increasing Access to the Results of Federally Funded Research," to ensure federally funded research is available to the public within one year of publication.
- NASA answered the mandate by creating an agency plan entitled "NASA Plan for Increasing Access to the Results of Scientific Research" and associated policy, NPD 2230.1, *Research Data and Publication Access*.

Principles in NASA SMD Strategic Plan for Scientific Data and Computing:

- Continued free and open access to scientific data for any use
- Improved ease of use and discoverability
- Enhanced science applications and new use cases
- Incorporates best practices and "state of the art" through partnerships

Earth Data and Systems are Evolving:

- Increasing archive and file sizes. More complicated data structures
- More user-friendly and data services
- What is the future direction?

Source: https://www.nasa.gov/open/researchaccess/public-access-results

Motivation (cont.)

Challenges in data access:

- "Decision Support Systems Analysts, the General Public, and University Undergraduates report the lowest levels of CSI" according to the 2017 CSI (Customer Survey Index). Over 50% of users.
- Surveys reveal that most non-professional users normally do not want to download and handle raw data as well as conduct heavy-duty data processing tasks.

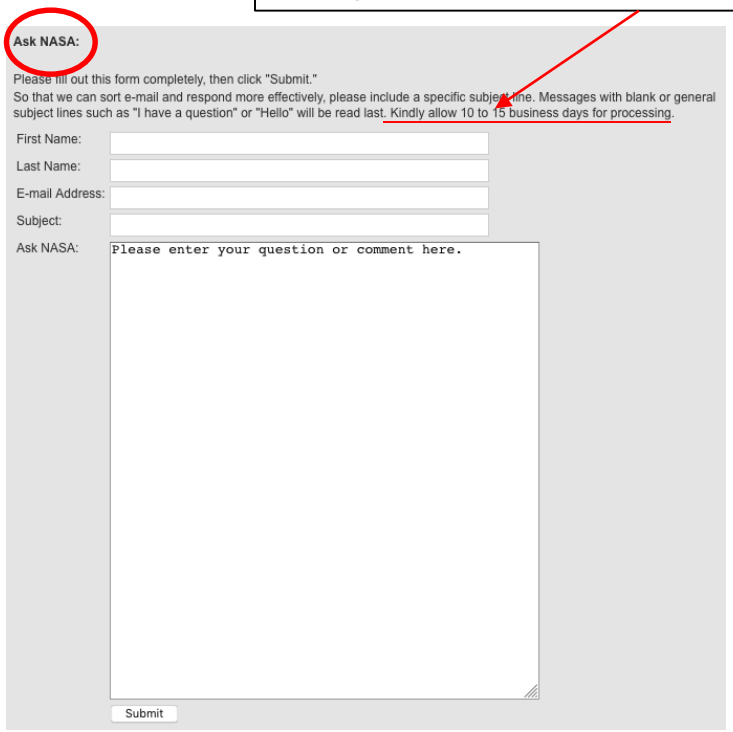
	2016	2016	2016	2017	2017	2017
	%	N	CSI	%	N	CSI
Type of User~						
General Public	14%	1,019	76	14%	1,037	76
Elementary, Middle, High School Teacher	1%	83	76	1%	86	77
University Professor	16%	1,129	80	16%	1,193	81
University Undergraduate Student	36%	2,550	76	9%	656	76
Other Education and Outreach	5%	349	79	5%	355	79
Earth Science Researcher	32%	2,304	79	32%	2,409	79
Earth Science Modeler	8%	574	78	9%	650	79
NASA-affiliated Scientist	2%	167	79	1%	102	80
Non-NASA-affiliated Scientist	4%	304	79	4%	320	78
NASA Science Team Member	7%	475	79	1%	68	80
Data Tool Developer/Provider	5%	359	77	5%	409	77
Decision Support Systems Analyst	5%	375	76	6%	429	76
University Graduate Student	0%	0	--	29%	2,204	77
Other User Type	8%	548	76	9%	656	77
Number of Respondents	**7,133**	**7,133**	**7,133**	**7,505**	**7,505**	**7,505**

2017 ACSI Survey Results

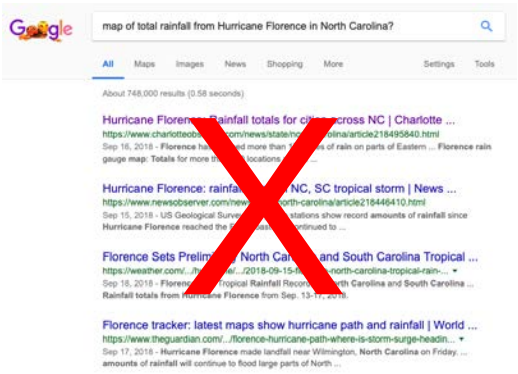
Source: the 2017 ACSI survey https://earthdata.nasa.gov/2017-acsi-survey-summary

Motivation (cont.)

Kindly allow 10-15 business days for processing

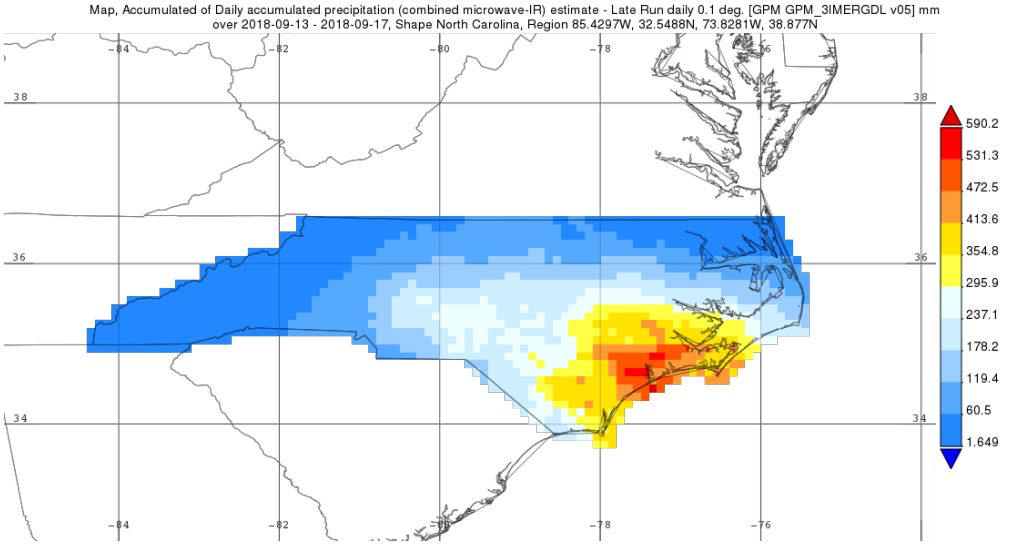


https://www.nasa.gov/about/contact/ask_nasa_form.html



Ask NASA

Map of total rainfall from Hurricane Florence in North Carolina?



Solution and Activities

Solution:

- Develop natural language processing (NLP) based data analysis and visualization infrastructure

Input (text, voice)

↓

Processing (text, voice)

↓

Analysis and Visualization

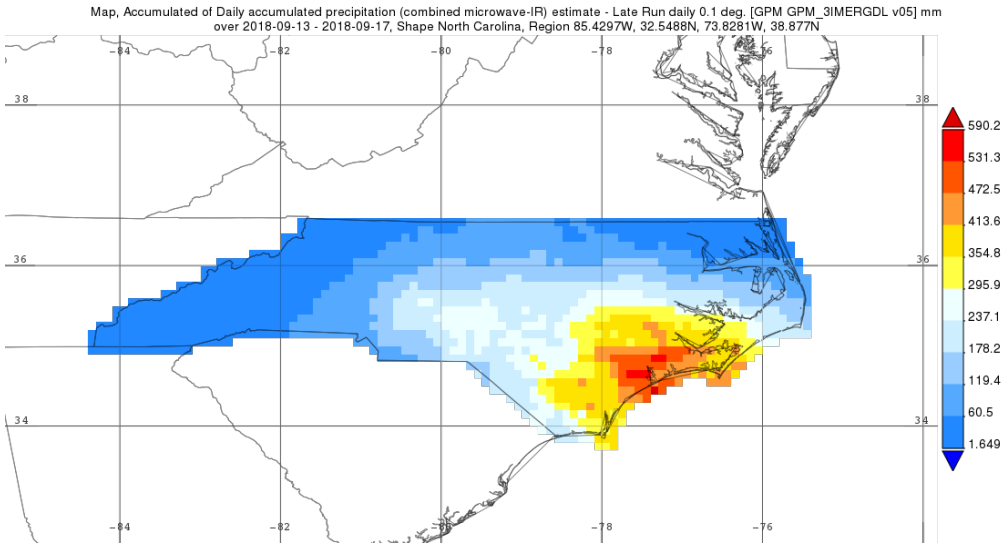
↓

Result (text, voice, graphics)

Ask NASA

What is the total rainfall from Hurricane Florence in North Carolina?

- *Collect user input info. about where, when, what, etc. (total rainfall map, Hurricane Florence, North Carolina)*
- *Call a backend system to process the inputs and generate the result (the rainfall map)*



Activities:

- Work with NLP experts at UMBC
- Use case development
- System design and prototyping

GES – DISC
Goddard Earth Sciences
Data Information Services Center

Summary

- Many challenges in Earth science data and information access for users at all levels
- NLP provides a simple (but difficult to develop) interface to ordinary users
- NLP provides inputs for backend processing (data analysis and visualization)
- Working with NLP experts to develop a prototype